

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/836,075ADATE: 02/04/98
TIME: 15:55:31

INPUT SET: S23159.raw

This Raw Listing contains the General
Information Section and up to the first 5 pages.

#8 1/2

SEQUENCE LISTING

ENTERED

1
2
3 (1) General Information:
4
5 (i) APPLICANT: MAERTENS, GEERT
6 STUYVER, LIEVEN
7
8 (ii) TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
9 AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
10 AGENTS
11
12 (iii) NUMBER OF SEQUENCES: 207
13
14 (iv) CORRESPONDENCE ADDRESS:
15 (A) ADDRESSEE: ARNOLD, WHITE & DURKEE
16 (B) STREET: P.O. BOX 4433
17 (C) CITY: HOUSTON
18 (D) STATE: TEXAS
19 (E) COUNTRY: USA
20 (F) ZIP: 77210-4433
21
22 (v) COMPUTER READABLE FORM:
23 (A) MEDIUM TYPE: Floppy disk
24 (B) COMPUTER: IBM PC compatible
25 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
26 (D) SOFTWARE: Microsoft Word 6.0 / ASCII text output
27
28 (vi) CURRENT APPLICATION DATA:
29 (A) APPLICATION NUMBER: 08/836,075
30 (B) FILING DATE: 21 Apr 1997
31
32 (vii) PRIOR APPLICATION DATA:
33 (A) APPLICATION NUMBER: PCT/EP95/04155
34 (B) FILING DATE: 23 Oct 1995
35
36 (viii) PRIOR APPLICATION DATA:
37 (A) APPLICATION NUMBER: EP 94870166.9
38 (B) FILING DATE: 21 Oct 1994
39
40 (ix) PRIOR APPLICATION DATA:
41 (A) APPLICATION NUMBER: EP 95870076.7
42 (B) FILING DATE: 28 Jun 1995
43
44 (ix) ATTORNEY/AGENT INFORMATION:
45 (A) NAME: KAMMERER, PATRICIA A.
46 (B) REGISTRATION NUMBER: 29,775

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/836,075ADATE: 02/04/98
TIME: 15:55:34

INPUT SET: S23159.raw

47 (C) REFERENCE/DOCKET NUMBER: INNS:004
48
49 (2) INFORMATION FOR SEQ ID NO: 1:
50
51 (i) SEQUENCE CHARACTERISTICS:
52 (A) LENGTH: 327 base pairs
53 (B) TYPE: nucleic acid
54 (C) STRANDEDNESS: single
55 (D) TOPOLOGY: linear
56
57 (ii) MOLECULE TYPE: cDNA
58
59 (iii) HYPOTHETICAL: NO
60
61 (iii) ANTI-SENSE: NO
62
63
64
65
66 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
67
68 ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCTCAK 60
69
70 GSGGTNNNNN NNCCGGGTGG CGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG 120
71
72 GGCCCCAGGN NGGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCACA ACCTCGTGGC 180
73
74 AGGCGACAGC CTATCCCCAA GGCTCGYCGG YCCGAGGGCA GGTCCCTGGGC TCAGCCCGGG 240
75
76 TATCCTTGGC CCCTCTATGG CAATGAGGGC TGCGGGTGGG CGGGNTGGCT CCTGTCCCCC 300
77
78 CGCGGCTCTC GGCCCAATTG GGGCCCC 327
79
80 (2) INFORMATION FOR SEQ ID NO: 2:
81
82 (i) SEQUENCE CHARACTERISTICS:
83 (A) LENGTH: 109 amino acids
84 (B) TYPE: amino acid
85 (D) TOPOLOGY: linear
86
87 (ii) MOLECULE TYPE: peptide
88
89
90
91 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
92
93 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
94 1 5 10 15
95
96 Arg Arg Pro Xaa Xaa Xaa Xaa Xaa Pro Gly Gly Gly Gln Ile Val Gly
97 20 25 30
98
99 Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Xaa Gly Val Arg Ala

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/836,075ADATE: 02/04/98
TIME: 15:55:37

INPUT SET: S23159.raw

	35	40	45
100			
101			
102	Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro		
103	50	55	60
104			
105	Ile Pro Lys Ala Xaa Arg Xaa Glu Gly Arg Ser Trp Ala Gln Pro Gly		
106	65	70	75 80
107			
108	Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Xaa Trp		
109		85	90 95
110			
111	Leu Leu Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro		
112	100	105	
113			

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 447 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

132	GACGGCGTGA ACTATGCAAC AGGGAAC TTG CCCGTTGCT CTTTCTCTAT CTTCCCTCTTG	60
133		
134	GCTTTGCTGT CCTGCTTGAC GGTTC CAACK ACCGCTCAG AGGTGCGCAA CGCATCCGGG	120
135		
136	GTGTATCATG TCACCAACGA CTGTTCCAAC TCGAGCATCA TCTATGAGAT GGACGGTATG	180
137		
138	ATCATGCACT ACCCAGGGTG CGTGCCCTGC GTTCGGGAGG ATAACCATCT CCGCTGCTGG	240
139		
140	ATGGCGCTCA CCCCCACGCT TGC GGTC AAA AAYGCTAGTG TCCCCACTRC GGCAATCCGA	300
141		
142	CGTCACGTCG ACTTGCTTGT TGGGGGNNCC ACGTTCTGTT CCGCTATGTA CGTGGGRGAC	360
143		
144	CTTTGCGGGT CTGTCTTCCT CGCTGGCCAG CTATTCACCT TTTCACCCCG CATGCACCAT	420
145		
146	ACAACGCAGG AGTGCAACTG CTCAATC	447

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 149 amino acids
- (B) TYPE: amino acid

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/836,075ADATE: 02/04/98
TIME: 15:55:40

INPUT SET: S23159.raw

153 (D) TOPOLOGY: linear

154

155 (ii) MOLECULE TYPE: peptide

156

157

158

159 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

160

161 Asp Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser
162 1 5 10 15

163

164 Ile Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Val Pro Xaa Thr Ala
165 20 25 30

166

167 His Glu Val Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp Cys
168 35 40 45

169

170 Ser Asn Ser Ser Ile Ile Tyr Glu Met Asp Gly Met Ile Met His Tyr
171 50 55 60

172

173 Pro Gly Cys Val Pro Cys Val Arg Glu Asp Asn His Leu Arg Cys Trp
174 65 70 75 80

175

176 Met Ala Leu Thr Pro Thr Leu Ala Val Lys Xaa Ala Ser Val Pro Thr
177 85 90 95

178

179 Xaa Ala Ile Arg Arg His Val Asp Leu Leu Val Gly Xaa Xaa Thr Phe
180 100 105 110

181

182 Cys Ser Ala Met Tyr Val Xaa Asp Leu Cys Gly Ser Val Phe Leu Ala
183 115 120 125

184

185 Gly Gln Leu Phe Thr Phe Ser Pro Arg Met His His Thr Thr Gln Glu
186 130 135 140

187

188 Cys Asn Cys Ser Ile

189 145

190

191 (2) INFORMATION FOR SEQ ID NO: 5:

192

193 (i) SEQUENCE CHARACTERISTICS:

194 (A) LENGTH: 327 base pairs

195 (B) TYPE: nucleic acid

196 (C) STRANDEDNESS: single

197 (D) TOPOLOGY: linear

198

199 (ii) MOLECULE TYPE: cDNA

200

201 (iii) HYPOTHETICAL: NO

202

203 (iii) ANTI-SENSE: NO

204

205

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/836,075ADATE: 02/04/98
TIME: 15:55:44

INPUT SET: S23159.raw

206
207 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
208
209 ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG 60
210
211 GACGTCAAGN TCCCGGGTGG TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG 120
212
213 GGCCCCAGGT TGGGTGTGCG CGCGACCAGG AAGACTTCCG AGCGGTCGCA GCCTCGTGAC 180
214
215 AGGCGACAGC CTATTCCTAA GGCTCGCCAG TCCGATGGCA GNNCCTGGGC TCAGCCAGGG 240
216
217 CATCCCTGGC CCCTCTATGG CAATGAGGGC TCGGATGGG CGGGATGGCT CCTGTCCCCC 300
218
219 CGCGGCTCTC GGCCCAGTTG GGGCCCC 327
220

221 (2) INFORMATION FOR SEQ ID NO: 6:
222

223 (i) SEQUENCE CHARACTERISTICS:

224 (A) LENGTH: 109 amino acids
225 (B) TYPE: amino acid
226 (D) TOPOLOGY: linear
227

228 (ii) MOLECULE TYPE: peptide
229
230
231

232 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
233

234 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
235 1 5 10 15
236
237 Arg Arg Pro Gln Asp Val Lys Xaa Pro Gly Gly Gly Gln Ile Val Gly
238 20 25 30
239
240 Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala
241 35 40 45
242
243 Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Asp Arg Arg Gln Pro
244 50 55 60
245
246 Ile Pro Lys Ala Arg Gln Ser Asp Gly Xaa Xaa Trp Ala Gln Pro Gly
247 65 70 75 80
248
249 His Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp
250 85 90 95
251
252 Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro
253 100 105
254

255 (2) INFORMATION FOR SEQ ID NO: 7:
256

257 (i) SEQUENCE CHARACTERISTICS:

258 (A) LENGTH: 447 base pairs

PAGE: 1

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION US/08/836,075A

DATE: 02/04/98
TIME: 15:55:47

INPUT SET: S23159.raw

Line

Error

Original Text